

WHAT IS CLAIMED IS:

1 1. A method for automatically indexing and retrieving a multimedia event,  
 2 comprising:  
 3 separating a multimedia data stream into audio, visual and text components;  
 4 segmenting the audio, visual and text components of the multimedia data stream  
 5 based on semantic differences, wherein frame-level features are extracted from the  
 6 segmented audio component are in a plurality of subbands;  
 7 identifying at least one target speaker using the audio and visual components;  
 8 identifying semantic boundaries of text for at least one of the identified target  
 9 speakers to generate semantically coherent text blocks;  
 10 generating a summary of multimedia content based on the audio, visual and text  
 11 components, the semantically coherent text blocks and the identified target speaker;  
 12 deriving a topic for each of the semantically coherent text blocks based on a set of  
 13 topic category models; and  
 14 generating a multimedia description of the multimedia event based on the  
 15 identified target speaker, the semantically coherent text blocks, the identified topic, and  
 16 the generated summary.

1 2. The method of claim 1, further comprising:  
 2 automatically identifying a hierarchy of multimedia content types.

1 3. The method of claim 2, wherein the multimedia content types include at least one  
 2 of speakers, anchors, interviews, correspondence reports, multimedia content segments,  
 3 general news stories, topical news stories, news summaries, and commercials.

1 4. The method of claim 1, further comprising:

2           converting the multimedia data stream from an analog multimedia data stream to a  
3 digital multimedia data stream; and  
4           compressing the digital multimedia data stream.

1    5.     The method of claim 1, wherein the extracted audio features from the audio  
2 component further comprise clip level features.

1    6.     The method of claim 1, wherein the multimedia event includes a news broadcast  
2 and the target speakers include news anchorpersons.

1    7.     The method of claim 1, wherein the step of identifying at least one speaker  
2 includes the process of identifying using Gaussian Mixture Models.

1    8.     The method of claim 1, wherein the generated multimedia description is  
2 represented by at least one of a text description, a video description and a story icon.

1    9.     The method of claim 1, further comprising:  
2           storing the generated multimedia descriptions in a database.

1    10.    The method of claim 1, further comprising:  
2           presenting the generated multimedia description to a user.

1    11.    The method of claim 10, further comprising:  
2           playing back the segment of the multimedia event corresponding to the generated  
3 multimedia description to the user.

1    12.    The method of claim 1, wherein the plurality of subbands comprises three

2 subbands.

1 13. The method of claim 12, wherein the frame level features in the three subbands  
2 are at least one of volume, zero crossing rate, pitch period, frequency centroid, frequency  
3 bandwidth and energy ratios.

1 14. A system that automatically indexes and retrieves a multimedia event, comprising:  
2 a multimedia data stream separation unit that separates a multimedia data stream  
3 into audio, visual and text components;

4 a data stream component segmentation unit that segments the audio, visual and  
5 text components of the multimedia data stream based on semantic differences;

6 a feature extraction unit that extracts audio features from the audio component and  
7 the audio features comprising a frame-level feature in a plurality of subbands;  
8 a target speaker detection unit that identifies at least one target speaker using the  
9 audio and visual components;

10 a content segmentation unit that identifies semantic boundaries of text for at least  
11 one of the identified target speakers, to generate semantically coherent text blocks;

12 a summary generator that generates a summary of multimedia content based on  
13 the audio, visual and text components, the semantically coherent text blocks and the  
14 identified target speaker;

15 a topic categorization unit that derives a topic for each of the semantically  
16 coherent text blocks based on a set of topic category models; and

17 a multimedia description generator that generates a multimedia description of the  
18 multimedia event based on the identified target speaker, the semantically coherent text

19 blocks, the identified topic and the generated summary.

1 15. The system of claim 14, wherein the multimedia description generator  
2 automatically identifies a hierarchy of multimedia content types.

1 16. The system of claim 15, wherein the multimedia content types include at least one  
2 of speakers, anchors, interviews, correspondence reports, multimedia content segments,  
3 general news stories, topical news stories, news summaries, and commercials.

1 17. The system of claim 14, further comprising:  
2 an analog-to-digital converter that converts the multimedia data stream from an  
3 analog multimedia data stream to a digital multimedia data stream; and  
4 a compression unit that compresses the digital multimedia data stream.

1 18. The system of claim 14, wherein the multimedia event includes a news broadcast  
2 and the target speakers include news anchorpersons.

1 19. The system of claim 14, wherein the target speaker detection unit identifies at  
2 least one target speaker using Gaussian Mixture Models.

1 20. The system of claim 14, wherein the multimedia description generator generates  
2 one or more multimedia description that are represented by at least one of a text  
3 description, a video description and a story icon.

1 21. The system of claim 14, further comprising:  
2 a database that stores the generated multimedia descriptions.

1 22. The system of claim 14, wherein the generated multimedia descriptions are  
2 retrieved from the database and presented to a user.

1 23. The system of claim 22, further comprising:  
2 a playback device that plays back the segment of the multimedia event  
3 corresponding to the generated multimedia description to the user.

1 24. The system of claim 14, wherein the plurality of subbands comprises three  
2 subbands.

1 25. A terminal that displays the multimedia descriptions generated by the multimedia  
2 description generator of claim 1.